Technical Report 1016

Personnel Enlistment Testing, Job Performance, and Cost: A Cost-Effectiveness Analysis

Dickie A. Harris, Rodney A. McCloy, John R. Dempsey, and Ani S. DiFazio
Human Resources Research Organization

Paul F. Hogan
Decision Science Consortium, Inc.

DTIC
ELECTE
JAN 23 1995
S G D

October 1994

19950118 065

DTIC QUALITY INSPECTED 3

**United States Army Research Institute
for the Behavioral and Social Sciences**

Approved for public release; distribution is unlimited.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.

1. AGENCY USE ONLY *(Leave blank)*

2. REPORT DATE
1994, October

3. REPORT TYPE AND DATES COVERED
Final Sep 91 - Nov 93

4. TITLE AND SUBTITLE
Personnel Enlistment Testing, Job Performance, and Cost:
A Cost-Effectiveness Analysis

5. FUNDING NUMBERS
MDA903-91-C-0242
63007A
792
1222
C1

6. AUTHOR(S)
Harris, Dickie A.; McCloy, Rodney A.; Demspey, John R.;
DiFazio, Ani S. (HumRRO); Hogan, Paul F. (DSC, Inc.)

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Human Resources Research Organization (HumRRO)
66 Canal Center Plaza, Suite 400
Alexandria, VA 22314

8. PERFORMING ORGANIZATION REPORT NUMBER
FR-PRD-93-35

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Research Institute for the Behavioral and
Social Sciences
ATTN: PERI-RS
5001 Eisenhower Avenue
Alexandria, VA 22333-5600

10. SPONSORING/MONITORING AGENCY REPORT NUMBER
ARI Technical Report
1016

11. SUPPLEMENTARY NOTES Task was performed under a Scientific Services Agreement issued by Decision Science Consortium, Inc. (DSC), 9302 Lee Highway, Suite 500, Fairfax, VA 22031. Contracting Officer's Representative, Peter Legree.

12a. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release;
distribution is unlimited.

12b. DISTRIBUTION CODE
--

13. ABSTRACT *(Maximum 200 words)*

The goals of this project were to (1) describe existing military selection and classification procedures, (2) formulate a set of alternative models, (3) develop an evaluation framework and associated criteria for comparing the cost-effectiveness of alternative models, and (4) assess the feasibility of the evaluation procedures. Previous reports addressed the first three goals. This report describes the pilot test of a cost-effectiveness model to evaluate alternative selection and classification models.

The Selection and Classification Evaluation Model (S&CEM) considered both desired level of performance and the costs of obtaining that performance goal. The S&CEM combined performance prediction equations with training, compensation, and recruiting costs. Next, a linear programming algorithm was used to solve for the most cost-effective mix of recruits that would meet the performance goal. The effectiveness and efficiency of a single-stage simultaneous selection and classification model were demonstrated by evaluating four test batteries. The value of each test battery was estimated as the cost necessary to meet a fixed performance goal. Strengths and weaknesses of the S&CEM are discussed.

14. SUBJECT TERMS
Selection
Classification
Job assignment
Recruiting costs
Cost-effectiveness
Linear programming algorithm
Training costs

15. NUMBER OF PAGES
50

16. PRICE CODE
--

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	Unlimited

NSN 7540-01-280-5500

Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std Z39-18
298-102

U.S. ARMY RESEARCH INSTITUTE FOR THE BEHAVIORAL AND SOCIAL SCIENCES

A Field Operating Agency Under the Jurisdiction of the Deputy Chief of Staff for Personnel

EDGAR M. JOHNSON
Director

Research accomplished under contract for
the Department of the Army

Human Resources Research Organization

Technical review by

Peter J. Legree
Jay Silva

Accesion For

NTIS CRA&I ☒
DTIC TAB ☐
Unannounced ☐
Justification ....

By ....
Distribution /

Availability Codes

Dist	Avail and / or Special
A-1	

NOTICES

DISTRIBUTION: Primary distribution of this report has been made by ARI. Please address correspondence concerning distribution of reports to: U.S. Army Research Institute for the Behavioral and Social Sciences, ATTN: PERI-POX, 5001 Eisenhower Ave., Alexandria, Virginia 22333-5600.

FINAL DISPOSITION: This report may be destroyed when it is no longer needed. Please do not return it to the U.S. Army Research Institute for the Behavioral and Social Sciences.

NOTE: The findings in this report are not to be construed as an official Department of the Army position, unless so designated by other authorized documents.

Technical Report 1016

Personnel Enlistment Testing, Job Performance, and Cost: A Cost-Effectiveness Analysis

Dickie A. Harris, Rodney A. McCloy, John R. Dempsey, and Ani S. DiFazio
Human Resources Research Organization

Paul F. Hogan
Decision Science Consortium, Inc.

Selection and Assignment Research Unit
Michael G. Rumsey, Chief

Manpower and Personnel Research Division
Zita M. Simutis, Director

U.S. Army Research Institute for the Behavioral and Social Sciences
5001 Eisenhower Avenue, Alexandria, Virginia 22333-5600

Office, Deputy Chief of Staff for Personnel
Department of the Army

October 1994

Army Project Number
2Q263007A792

Manpower and Personnel

Approved for public release; distribution is unlimited.

FOREWORD

The U.S. Army has embarked on a line of research to evaluate and improve its existing selection and classification system. Toward this goal, the Selection and Assignment Research Unit (SARU) of the Manpower and Personnel Research Division (MPRD) at the U.S. Army Research Institute for the Behavioral and Social Sciences (ARI) contracted with the Human Resources Research Organization to identify and evaluate alternative selection and classification models. As part of this contract, this report presents both an exposition of the methodological framework for evaluating selection and classification models, and an application of this framework.

EDGAR M. JOHNSON
Director

ACKNOWLEDGMENTS

The authors would like to recognize the contributions of numerous colleagues who contributed to this report. The Contracting Officer's Technical Representative for this effort was Peter Legree of the U.S. Army Research Institute for the Behavioral and Social Sciences. The authors would like to thank Teresa Russell, Janice Laurence, and Mary Ann Statman for contributing their time to read and comment on numerous drafts of the report. We are also indebted to the many Service representatives who took time to answer our questions about selection and classification within their respective Service.

PERSONNEL ENLISTMENT TESTING, JOB PERFORMANCE, AND COST: A COST-EFFECTIVENESS ANALYSIS

EXECUTIVE SUMMARY

Requirement:

The Army strives toward efficient personnel selection and classification methods. Although considerable progress has been made over the years, the more the Army can learn about the costs and benefits of alternative selection and classification methods, the more effective its personnel management systems can be. The goals of the Selection and Classification Models project were to (1) describe existing military selection and classification procedures, (2) formulate a set of alternative models, (3) develop an evaluation framework and associated criteria for comparing the cost-effectiveness of alternative models, and (4) assess the feasibility of the evaluation procedures. Previous reports addressed the first three goals. This report describes the pilot test of a Selection and Classification Evaluation Model (S&CEM).

Procedure:

A cost-effectiveness approach that considers both the desired level of performance and the costs of obtaining that performance goal was employed to evaluate the efficiency of alternative test batteries for selection and classification. A linear programming (LP) model was used to estimate the cost-effectiveness of the batteries by simulating a one-stage simultaneous selection and classification process. This framework utilized performance prediction equations for nine occupational areas computed from a given battery, along with training, compensation, and recruiting costs, and solved for the most cost-effective mix of recruits that met the performance goals for each job family. Data were obtained from the Project A database to evaluate four test batteries. Battery A was the Armed Forces Qualification Test (AFQT). Battery B contained the verbal, quantitative, technical, and speed composites of the Armed Services Vocational Aptitude Battery (ASVAB). Battery C added a spatial composite to the ASVAB and Battery F added ABLE, a measure of the willingness to perform, to Battery C. The potential value of improved testing (e.g., Battery A versus Battery C) was estimated as the reduction in total cost necessary to meet the established performance goals for all jobs.

Findings:

The LP cost estimates suggested that adding a spatial composite to the ASVAB may save up to $114 million in recruiting, training, and compensation costs for an Army recruit cohort over four years. The results also indicated that the spatial composite would be particularly useful in finding occupational areas where lower quality recruits (i.e., AFQT Category IIIB and IV) with above average spatial ability would perform well. Including ABLE in an enlistment test battery was estimated to save an additional $160 million relative to Battery C. However, a higher quality mix of recruits was chosen when the information provided by ABLE was used to make selection and classification decisions. This pilot test confirmed the potential of the LP method, within the context of a cost-effectiveness framework, to provide relatively clear answers to questions about the relative value of alternative selection and classification batteries.

These savings estimates should not be considered as absolute values given that validities were obtained from more or less ideal experimental conditions. Further, the "savings" do not consider the developmental and implementation costs of the additional/alternative measures. However, the S&CEM is a useful tool for examining alternative selection and classification batteries in terms of their cost-effectiveness.

Utilization of Findings:

The methods developed and tested in connection with this phase of the research effort were used to assess the effectiveness and efficiency of alternative enlistment test batteries. The evaluation framework can be applied to a number of different policy issues facing the Army. Examples of some specific policy questions and issues that may be evaluated with the current framework include

(1) How would results change if we include more realistic factors, such as applicant preferences and training seat availability, directly in the simulations? What is the value (cost) of limiting (expanding) applicant choices in classification?

(2) What are the expected costs associated with eliminating a test, such as Numerical Operations, from the current selection and classification battery?

(3) What is the "optimal" set of tests to include in an aptitude battery? Can an "optimal" battery be constructed using the framework?

(4) What is the dollar value of the tradeoff between tests with less adverse impact, but less predictive precision?

PERSONNEL ENLISTMENT TESTING, JOB PERFORMANCE, AND COST: A COST-EFFECTIVENESS ANALYSIS

CONTENTS

Page

INTRODUCTION 1

- Objectives 1
- Previous Research on the Value of Selection and Classification 3
- Overview of the Cost-Effectiveness Method 9

SELECTION AND CLASSIFICATION EVALUATION MODEL 11

- Setup of the Problem 11
- Objective Function 11
- Model Dimensions and Data 12
- Major Components 14

METHOD AND RESULTS 27

- Generation of the Synthetic Sample 27
- Prediction Equations 29
- Results 33

DISCUSSION AND CONCLUSIONS 35

REFERENCES 39

LIST OF TABLES

Table 1. The Military Occupational Specialties (MOS) Studied in Project A 16

2. Composite Scores From the Project A Predictor Battery 17

3. Job Performance Criterion Measures Used in Project A Concurrent Validation Samples 18

CONTENTS (Continued)

Page

Table 4. Frequency Distribution of Residuals 28

5. Multiple Correlations for 12 Alternative Equations for Predicting Performance 31

6. Multiple Correlations for Six Alternative Selection and Classification Batteries 32

7. Selection and Classification Batteries Considered 33

8. Results from Selection/Classification Battery Evaluation 34

LIST OF FIGURES

Figure 1. Cost-effectiveness comparisons of personnel selection and classification tests 34

PERSONNEL ENLISTMENT TESTING, JOB PERFORMANCE, AND COST: A COST-EFFECTIVENESS ANALYSIS

I. Introduction

What is the value to the Army of additional selection and classification tests? What is the cost, if any, of eliminating some of the tests that are now given to Army applicants? These are practical questions that have plagued researchers as well as policy makers in the military testing field and, more broadly, in the aptitude testing and entrance screening applications of industrial psychology and psychometrics. Answers to these questions will help the Army determine the resources that should be allocated to selection and classification testing in general. A framework that provides estimates of the payoff of selection and classification testing, in terms of savings in real budget expenditures, will help the Army determine which tests have low payoffs and, perhaps, should be eliminated. It will also help to focus testing research and development on those areas where the returns are apt to be the highest.

In this report, we present both an exposition of the methodological framework for evaluating selection and classification tests, and an application of this framework. The cost-effectiveness framework we developed to answer the questions raised in the opening paragraph permits us to estimate the value of selection and classification testing in terms of the dollar cost of recruiting, training, and compensation resources necessary to obtain a first-term enlisted force of a desired capability or expected performance level. In this framework, better selection and classification tests affect these expenditures by (a) screening out applicants who are not likely to provide a cost-effective contribution to first-term readiness of the force and (b) determining the best match of an applicant's aptitudes with the demands of the occupation, so that the best use is made of the soldier's talents. We apply this framework to four progressively complex batteries of selection and classification tests, and obtain the incremental value, in terms of recruiting, compensation, and training expenditures saved, of the information provided by additional testing.

Objectives

The more the Army can learn about the costs and benefits of alternative selection and classification methods, the more effective its personnel management systems can be. Ideally, a simulation would permit evaluation of a full range of "what if" questions focused on the effects of changes in (a) labor supply, (b) recruiting procedures, (c) selection and classification measures, (d) decision-making algorithms, (e) applicant preferences, (f) various organizational constraints, and (g) changing organizational missions on such things as (1) the distribution of individual performance in each job, (2) attrition, (3) discipline problems, and (4) morale. Further, it would be desirable to have a good estimate of the specific costs involved in each change.

Though a comprehensive "what if" capability is not possible currently, the Army is now in a good position to take major steps toward such a personnel management capability. The Project A database (Campbell & Zook, 1990) makes it possible to begin exploring the limits of the gain that classification can provide compared to random assignment. This database provides (1) a full range of criterion variables that can be used to model alternative selection and classification goals, (2) an extensive battery of new tests that sample a broad range of different predictor domains, and (3) a sample of jobs chosen to represent the full range of Military Occupational Specialties (MOS) in the Army. Project A and the Linkage Project (Harris et al., 1991; McCloy et al., 1992) provide the rudiments of a capability to answer questions about the costs and benefits of alternative selection and classification models. Accordingly, the current project had the following objectives:

(1) Describe the existing selection and classification procedures of the Army, Navy, Air Force, and Marines, documenting all decision points, the information used at each, the constraints that operate, etc.

(2) Formulate a set of selection and classification models using existing databases, research results, and organizational policy.

(3) Develop criteria and an evaluation framework to compare the costs and effectiveness of the alternative models.

(4) Pilot test the feasibility of the evaluation framework.

Three of the objectives have been accomplished. Laurence and Hoffman (1993) described existing selection and classification procedures and formulated a set of alternative selection and classification models. Hogan, McCloy, Harris, and McWhite (1993) detailed the criteria and a framework to evaluate the cost-effectiveness of the alternative selection and classification models. This report describes the pilot test of the evaluation framework.

The pilot test had two purposes. The first purpose was to develop and test the Selection and Classification Evaluation Model (S&CEM). The second was to evaluate alternative sets of selection and classification methods. There were two key issues that shaped the structure of the cost-effectiveness model. First, the model had to estimate the potential **dollar value** to the Army of improved selection and classification methods. Second, the model had to be sufficiently flexible to consider both single-stage and multi-stage selection and classification systems. Using this design strategy, the savings from more intensive testing of an already selected group could be estimated.

Hogan et al. (1993) described four general selection and classification models. One model was the present selection and classification system used by the Army. In this model, most applicants are sent to a Military Examination Processing Station (MEPS) or Mobile Examination Team (MET) site where they complete the Armed Services Vocational Aptitude Battery (ASVAB) for the record. Their scores on the ASVAB (e.g., the Armed Forces Qualification Test (AFQT) and the Aptitude Area (AA) composite scores) are used to select and classify the applicants to specific Military Occupational Specialties (MOS). The AFQT score serves as the principal selection measure. The classification decisions are based on the various AA composite scores. Each MOS has certain AA composite "cut" scores that must be exceeded by the recruit for him or her to be eligible. In practice, the AA "cut" scores are set such that most recruits qualify for all MOS. The Army's system is a two-stage model because the selection and classification processes are independent.

A two-stage model differs from a single-stage model in which selection and classification occur simultaneously. The major distinction between the two models is that the classification process in a single-stage procedure at least partially determines the nature of the selected group, i.e., the recruit quality mix, and therefore, also impacts on recruiting costs. This is not true in a two-stage model, where the recruiting costs are solely a function of the selection strategy. Classification, in this case, means assigning a predetermined pool of applicants to jobs. Since the cost-effectiveness model developed in this study was designed to measure system efficiency in terms of the recruiting, training, and compensation costs necessary to meet performance goals, a single-stage model was examined in the pilot test.

The operational distinction between selection and classification occuring in a single-stage, or in two discrete stages, is whether the process of classification affects the nature of the population being classified, or whether this population is held constant. If it is held constant, it is determined by selection. Recruiting costs are no longer relevant, because they have been determined by the selection stage. On the other hand, if the classification also affects the distribution of entrants (i.e., single-stage model), recruiting costs must be considered. In particular, if classification takes into account a particular applicant characteristic (e.g., AFQT score) and determines the number of entrants with that characteristic, then the costs of increasing (or decreasing) the number of entrants with that characteristic must be considered. That is, the supply conditions of that characteristic must be considered. Since one of the objectives of the pilot test was to estimate the potential dollar value of improved selection and classification information we wanted to estimate the total cost (i.e., recruiting, training, and compensation) associated with selection and classification. Thus, the pilot test of the S&CEM was conducted using a single-stage (i.e., simultaneous selection and classification) variant of the Army's two-stage system (Hogan et al., 1993).

This report is organized as follows. Chapter II outlines the development and pilot testing of the S&CEM. Chapter III describes the results of the cost-effectiveness analyses of four selection and classification batteries. A discussion of the implications of the cost-effectiveness evaluation is presented in Chapter IV, along with policy ramifications and approaches for mitigating potential weaknesses through additional research and development. To put the S&CEM in perspective, the next section presents a brief discussion of previous efforts to evaluate selection and classification methods.

Previous Research on the Value of Selection and Classification

Existing criteria for evaluating testing methods stem largely from Brogden (1946, 1949). This seminal model, and contributions that followed in the same or similar spirit (most notably, Cronbach & Gleser, 1965; Hunter & Schmidt, 1982), focused largely on the selection criterion for a single job emphasizing the statistical relationship between predictor variables and the criterion or outcome variable. The stronger the statistical relationship between the predictors and the criterion or outcome variable, the better or more valuable the particular predictor or set of predictors is judged as a screening or classification tool.

A necessary condition for the efficacy of any selection or screening method is that its prediction of performance (conditional on the predictor or predictors) improves upon an unconditional prediction or the expected outcome under a random hiring or assignment policy. Whatever value the screening method may have will be a monotonic function of (and in some cases proportional to) the ability to improve upon the unconditional prediction of performance. This, perhaps, explains the initial focus on the statistical relationship between screening variables and performance as the primary criterion from which to judge testing methods. However, it became apparent that this was not a sufficient criterion. A statistical measure, such as the coefficient of determination or validity coefficient, does not address the economic value of selection (or classification). Thus, the focus has shifted toward the net benefits, in dollar terms, of a given selection (and/or classification) method to an employer.

We consider, first, criteria for evaluating the benefits of selection for a single job. Given the fundamental ideas from this literature, we will then consider "classification"--assignment of individuals across jobs based on differences in aptitudes and expected performance, and finally, simultaneous selection and classification decisions. This literature is also reviewed in somewhat greater detail in Zeidner and Johnson (1989).

In these models, the concept of selection "utility" is derived as follows. Let y_i be the dollar value of output or performance of the ith individual. Then, we can estimate the relationship between y_i and a predictor variable, such as the individual's score on an aptitude test (X_i), through the linear regression

$$y_i = \alpha + \beta X_i + \mu_i \tag{1}$$

where α is a constant and β is the slope coefficient of the predictor, X_i. In this exposition, y_i is the dollar value of the output, or performance metric, for individual i, and μ_i is a residual with $\mu_i \sim N(0,\sigma^2)$.

In this equation, $\alpha = Y^* - \beta X^*$, where "*" denotes the sample mean of the variable. Random selection of applicants implies that the average test score is X^* and average performance is Y^*. The increase in value or "utility" from setting a "cut" score for X, such that the mean value of X for those offered (and accepting) the job is X', is given by

$$\Delta U = N\,[\,(\alpha + \beta X') - (\alpha + \beta X^*)\,] \tag{2}$$

which is equal to

$$\Delta U = N \beta X' \tag{3}$$

when X is measured as a Z score based on the applicant population. The more readily recognized equation is obtained by noting that $\beta = \Sigma y_i x_i / (\Sigma x^2_i)$ when X and Y are measured as deviations from the mean. This is equal to $r_{xy}\sigma_y/\sigma_x$, where r_{xy} is the correlation coefficient between x and y, and "σ" denotes the standard deviation. If X is measured in standard normal form, $\sigma_x=1$; hence

$$\Delta U = N r_{xy} \sigma_y X' \tag{4}$$

where σ_y, or SD_y as it is denoted in much of the literature, is the dollar-valued standard deviation in performance. The dollar increase in utility associated with the selection of an applicant with mean predictor X' is the above expression divided by N, the number of entrants.

The equation derived above is the fundamental relationship used to describe the economic value or benefits of selection. In practice, the criterion variable, y, is a physical measure of on-the-job performance, and not a dollar measure. Dollar values enter the equation through σ_y, the standard deviation in individual performance. Attempts have been made to estimate the dollar value to the employer of a standard deviation in individual performance either through subjective estimation (expert judgment) or cost accounting methods (Hunter & Schmidt, 1982).

The above model employs only a single predictor or explanatory variable. The model itself can be expanded as a multivariate regression model, with k explanatory or predictor variables. The form of the regression model is

$$Py_i = \sum_{j=1}^{k} X_{ij}\,\beta_j + u_i \tag{5}$$

where P is the dollar value of a physical unit of performance, y_i; the X_{ij} are the characteristics of the applicant, which may include test scores but may also include other characteristics that are related to performance; and

$$\frac{\partial Py_i}{\partial X_{ij}} = \beta_j \tag{6}$$

where β_j is the dollar value of the change in performance when the characteristic X_j increases. In this equation, the value of selection for a new entrant cohort of size N is given by

$$N*E\,[Py_i - Py^*] = N*\sum_{j=1}^{k} (X^g{}_j - X^*{}_j)\;\beta_j \tag{7}$$

where X^*_j is the mean of characteristic X_j for a randomly selected applicant group, and X^g_j is the mean of characteristic j for the group selected on the basis of predicted performance.

This equation is equivalent to the original Brogden equation in both the univariate and multivariate case. The dollar value of performance, P, is multiplied by the measure of output or performance in physical units, y, prior to estimating the multivariate regression. Then, the coefficients (i.e., the β_j's) are interpreted as the marginal (dollar) value of characteristic X_j in producing the value of performance, Py. The net dollar value of the selected group compared to the random group is then given by the expected value of the difference in performance between the selected group and the random entrants. The net value of selection is equal to the gross value, from the equations above, less the cost of developing and applying any selection tests and/or the costs of collecting other information used for applicant screening.

There are several shortcomings associated with this simple model of the value of selection. Some of these can be addressed by expanding the model. However, some conceptual difficulties remain, particularly when the model is expanded to the public sector.

The simple model fails to explain why entry level selection is needed to obtain the benefits of a better-than-random distribution of worker performance. One alternative is simply to let all applicants enter the organization, observe their actual on-the-job performance for a period sufficient to provide a reasonable estimate of individual productivity, and selectively retain the best workers. For entry-level screening to be optimal, there must be costs associated with this procedure that are reduced through screening. Obvious costs include: initial hiring or recruiting costs, the costs of entry-level firm-specific training, any "damage" costs that can be imposed on the employer by poorly performing new employees prior to on-the-job observation of their performance, and costs of monitoring or detecting actual performance of the recent hires.

Another shortcoming of the basic model is that it does not account for the costs associated with obtaining new entrants. This takes the model outside of a traditional decision-theoretic framework because it implies a zero cost to "type II" errors--rejecting applicants who would have performed well. In the model, a "cut" score (in terms of X) is set and a distribution of employees with a mean predictor score above the "cut" score emerges. The best that can be said is that, implicitly, this distribution of willing applicants with predictor scores at or above the "cut" score is exogenous, perhaps reflecting a constant wage offer and a fixed amount of resources devoted to advertising and other factors that may affect this distribution. However, if applicants with higher predictor scores are more valuable to the organization, more resources will be devoted to attracting them, which will increase the supply. An equilibrium should be reached where the marginal recruiting costs are just equal to the marginal (expected) benefits of the higher scoring recruits.

The basic model, however, does not include an explicit supply curve of applicants of varying potential, as measured by X_i. Instead, the distribution is apparently fixed. In a decision-theoretic framework, the cost of raising the "cut" score and rejecting some applicants with low values of X_i who would have performed well is higher recruiting costs associated with obtaining the organization's workforce from a smaller population. If, however, the distribution of willing applicants by predictor score, X, is fixed or exogenous, this is not part of the decision process. Instead, one simply goes down the distribution of X's, starting from the highest, until N acceptances are obtained.

The pool of applicants can become endogenous. For example, by making the entry-level wage or recruiting expenditures part of the selection decision, one can increase the number of applicants and be more selective. This increase in recruiting costs should be balanced with the value of the increase in expected performance. By making the applicant pool a function of choices regarding entry wages and recruiting expenditures, the costs of rejecting applicants who would have been adequate performers is taken into account in the higher recruiting and entry wage costs that result.

If an employer has many jobs but only a specific, non-overlapping population applies for each type of job, or if all marginal costs and marginal products are the same and independent of specific jobs, then there is no operational distinction between selection and classification.[1] Conceptually, however, when there is more than one type of job to be filled in the organization, one can consider the general case of selection and classification as two distinct decisions: offering applicants employment in general (selection) and assigning them to a particular type of job within the organization (classification). If one makes this conceptual distinction, then the criteria for classification efficiency focus on the assignment of a given number of new hires to particular jobs, conditional on selection.[2]

In Brogden's (1951) model, which incorporated multiple jobs, individuals were assigned to the job for which the criterion score was the greatest. This criterion for classification, which Zeidner and Johnson (1989) call maximization of mean predicted performance (MPP), has also been considered the "optimal" assignment policy:

> Optimal assignment of all selected personnel could be accomplished, without considering constraints, by assigning each recruit to the job family corresponding to his highest test composite score, thus providing the largest MPP score obtainable for a specified set of assignment variables and sample of individuals (Zeidner & Johnson, 1989, p. 1-18).

Given this definition of "optimal assignment," the criteria used to value the benefits of classification have generally evolved from the original work of Brogden. Zeidner and Johnson (1989), following Hunter and Schmidt (1982), noted that the assumption that all jobs are of equal value is undoubtedly false. Hence, some effort should be made to assign different values or importance weights to different jobs. Optimal assignment then attempts to maximize MPP, weighted by these job valuation factors, in a "hierarchical" model of job assignment.

Estimation of the net benefits of classification is made with respect to an alternative policy of random assignment.[3] The benefits of classification, compared to random assignment, can be estimated using Brogden's dollar value of the standard deviation in performance, SD_y, in much the same way as it is done for a single job.

---

[1] It is important that an individual's expected performance vary across jobs. However, contrary to some statements in the literature, there is still a classification problem even if an individual's performance is not predicted to vary across jobs. If the criterion for "optimal" classification is the maximization of mean predicted performance (MPP), then the performance of a fixed pool of applicants is independent of assignments if an individual's expected performance is independent of the assignment. However, if training costs vary differentially across jobs and individual performance is correlated with training costs, then under a more complete definition of "optimal" classification, assignment will make a difference.

[2] In principle, one can consider three possible sequences: (a) selection then classification, (b) classification then selection, and (c) concurrent selection and classification. In general, concurrent selection and classification will be more efficient because it simultaneously considers all the costs (i.e., recruiting, training, and compensation) and benefits associated with a personnel decision.

[3] Because the performance models in this literature are typically linear, random assignment is equivalent to assuming that performance is measured as the mean for the sample (i.e., in expected value terms).

There are several problems with this estimate of the net value of classification. First, it does not consider training costs, recruiting costs,[4] or other costs associated with the personnel system that can be affected by the allocation of individuals across jobs. Second, when training and other costs enter the classification decision, the classification rule should become that of classifying to maximize net benefits, not mean predicted performance. Net benefits include the estimated value of performance, perhaps using a variant of Brogden's equation, less the costs of generating that performance. In many instances, it is likely that training costs (perhaps through the costs of premature attrition) as well as other costs will vary with the allocation decisions made. If so, it will no longer be the case that the "optimal" assignment is necessarily the one that maximizes MPP. In particular, individuals may not be allocated to jobs for which their predicted performance is highest, but to jobs for which their contribution to net benefits is greatest.[5] Further, the problems with estimating a dollar value for performance are now compounded somewhat by the problems associated with placing relative values or importance weights across jobs.

Finally, for theoretically "optimal" selection and classification decisions, these processes should be conducted simultaneously, not sequentially.[6] The reason for this is that the best criterion for selection and classification is the net benefits of the resulting job match. The net benefits are the value of the performance expected to be generated in the job by the match, less the costs (e.g., recruiting and training costs) of achieving the match. Hence, the selection criteria should be related directly to the classification criteria. Moreover, the pool of applicants should be endogenous for joint selection and classification. Recruiting costs and initial wage offers should be part of the policy variables and costs used to

---

[4] Recruiting costs are relevant only if classification affects selection, or if selection and classification are simultaneous. In the more narrow problem of assigning a fixed number of new recruits to jobs, recruiting costs are not relevant (i.e., they are sunk costs).

[5] As an illustration, consider a case with two classes of employees and two types of jobs. Training costs and expected performance values are shown in the following table:

Classification Decision				
Employee	Job 1		Job 2	
	Training	Performance	Training	Performance
A	50	100	60	110
B	80	90	70	90

If individual A is allocated to Job 1, and B to Job 2, net benefits (performance value less training costs) are $70. If we make the opposite allocation, net benefits are $60. However, to maximize the value of performance, or MPP, A would go to Job 2.

[6] An exception to this is if there is reason to economize on classification testing. For example, suppose there is a classification test that is very costly to administer. Clearly, some less costly forms of screening or selection should be conducted, with the more expensive tests administered only to those more likely to be ultimately selected.

affect the applicant pool from which selection and classification decisions are made.[7]

Overview of the Cost-Effectiveness Method

As seen in the previous section, evaluation of the benefits of a selection and classification program for hiring new employees has evolved from the calculation of a very narrow statistical index to a comprehensive analysis of the effects of a selection and classification system on both the performance of and cost to the organization. Evaluation methods should consider selection and classification as part of a personnel management system, and include not only the effects on the expected performance of new employees, but also the effects on the costs of recruiting and training employees, the costs associated with premature attrition, and the total costs to the organization.

Two general approaches to evaluating selection and classification methods that take costs into account have been documented in the literature. The following sections discuss these frameworks.

Net Benefit Criterion

In the first approach, originating with Brogden (1946) and applied most comprehensively to the Army by Nord and Kearl (1990), an attempt is made to calculate the net benefit of selection and classification. To do this, one must compare benefits and costs in a common metric--typically dollars. The models are compared based on their net benefit--the value of the expected performance resulting from the model less the cost generated to produce that expected performance. That is, alternative selection and classification models are ranked based on

$$Net\ Benefit_j = V(P_j) - C_j(P_j) - SC_j(P_j) \tag{8}$$

where the net benefits associated with alternative j are the value of performance produced under alternative j, $V(P_j)$, less the costs of producing that level of performance, $C_j(P_j)$, and the costs associated with the selection and classification process under alternative j, $SC_j(P_j)$. V(...) is the valuation function for performance, $C_j(\ldots)$ is a cost function for producing the level of expected performance, P_j, and $SC_j(\ldots)$ is the cost function for the selection and classification testing procedures.

Conceptually, this cost-benefit criterion is sound. In practice, however, it is difficult to specify the valuation function, V(...), which places a dollar value on performance. Note that there are potentially two valuation problems: (a) the valuation of performance for a given job; and (b) the valuation of performance across jobs. The Brogden approach is one way to attempt the former. Ultimately, however, this valuation becomes subjective. We have argued (Hogan, et al., 1993) that it is better to avoid attempting to place a dollar value on performance, if possible.

---

[7] An intuitive proof, based on Le Chatelier's Principle, is that permitting the expansion of the applicant pool (by incurring additional recruiting costs) and setting selection criteria (by looking ahead to classification decisions) increases the degrees of freedom over which one can optimize. Fixing either of these reduces flexibility and, therefore, must result in the same or a lower level of net benefits.

Cost-Effectiveness Criterion

The cost-effectiveness approach holds some measure of desired performance constant and compares the costs of alternative ways of achieving that goal. This is the approach taken in the Rand model (Fernandez & Garfinkle, 1985), the "opportunity cost" model of Nord and Schmitz (1989), and the McCloy et al., (1992) accession quality Cost-Performance Tradeoff Model (CPTM). The value of alternative selection and classification methods is the cost savings, relative to some baseline in achieving the desired level of performance.

This approach avoids the difficult problem of placing a dollar value on expected performance by comparing all alternatives at the same level of performance. Less efficient selection and classification models will produce this level of performance only at a higher total cost. Hence, the benefits of a given alternative are measured as the difference between costs of an alternative and costs of the baseline.

Because the level of performance is held constant, the value of performance is also constant across alternatives. Hence, to compare alternative i with alternative j, we have

$$\begin{aligned} NetBenefit_j - NetBenefit_i = {} & V(P) - C_j(P) - SC_j(P) \\ & - [V(P) - C_i(P) - SC_i(P)] \end{aligned} \qquad \textbf{(9)}$$

Because P and V(P) are constant across the alternatives, we have

$$NetBenefit_j - NetBenefit_i = C_i(P) + SC_i(P) - C_j(P) - SC_j(P) \qquad \textbf{(10)}$$

Using this equation, alternative models can be ranked based on their cost-savings relative to a baseline case.

What is lost in the cost-effectiveness formulation is the ability to compare alternatives that provide different levels of performance or benefits at different costs. In practice, we do not believe this is a significant limitation. Trained, ready, first-term personnel are important components of the Army's process for producing combat capability. The level of first-term performance required is derived from the overall Army plan. For the most part, any model of selection and classification adopted by the Army would be required to produce about the same level of performance in the first-term force.

Using either criterion, evaluation of the alternative requires measuring the effect on the ability to predict performance, based on the information available under the alternative. Performance equations have to be estimated that predict an applicant's performance across occupations, conditional on the information available. This permits the eventual simulation of the effects on selection and classification decisions. The next chapter describes such a simulation model.

II. Selection and Classification Evaluation Model

This chapter describes the Selection and Classification Evaluation Model (S&CEM) in greater detail. The topics discussed are: (1) the setup of the problem, (2) the objective function, (3) the dimensions of the model, and (4) the components of the S&CEM.

Setup of the Problem

Recruiters are actively employed in developing and pursuing leads concerning potentially qualified people who might be willing to enter the Army. Under the current selection process, an applicant's qualifications for military service are generally summarized by his or her scores on the Armed Forces Qualification Test (AFQT) and education credential. The key educational distinction is whether or not the applicant has graduated from high school. An applicant's score on the AFQT indicates his or her aptitude for the occupations offered by the Services. Applicant scores are typically summarized by one of six discrete categories: I, II, IIIA, IIIB, IV, and V. Category I is the highest, whereas Category V recruits--those who score in the lowest decile on the AFQT--are prohibited by law from entering service.

Applicants who are willing to serve, and who qualify under the current criteria for enlistment, enter the military for a specified term of service. An individual recruit's relative performance may vary across occupations, and an important consideration of the Army is to place the right recruit in the right occupation. During the first year of service, the recruit receives basic training and, in most instances, initial skill training. The recruit may not have the perseverance or ability to complete training, and may leave the service prior to completion. Upon successful completion of training, the recruit is assigned to a unit. His or her performance in that unit jointly produces military readiness and on-the-job training.

The problem for the Army, as we frame it, is to choose the number and quality mix of recruits (selection). Further, recruits must be allocated across occupational groups to meet first-term performance goals at the lowest cost.

Objective Function

The objective function of the linear programming (LP) model is to choose the number of accessions from each recruit category, defined by scores on selection and classification tests, to minimize the present value of the costs of achieving a given level of performance, by occupation, over the first-term of service. Recruits contribute to the performance constraint in an occupation as they progress through the system, but recruiting, training, and compensation costs are also incurred. Performance varies by recruit category and by occupation within recruit category.

Expected performance over the first-term of service is calculated for each recruit category (j) and occupation (i) as

$$P^*_{ij} = \sum_{t=1}^{48} S^t_{ij} P_{ij} \qquad \textbf{(11)}$$

where P^*_{ij} is expected performance of a recruit from category j in occupation i over the first-term of service, S^t_{ij} is the probability of a recruit from category j in occupation i surviving to month of service t, and P_{ij} is the expected performance from a recruit in category j, occupation i.

Similarly, we can calculate T^*_{ij} and C^*_{ij}, expected training costs and expected compensation costs for recruits in category j, occupation i, respectively, over the first-term of service discounted to the entry point. Then, the model chooses the number of recruits from recruit category j allocated to occupation i, A_{ij}, to minimize the costs subject to meeting performance goals, P'_i, for each occupation.

Formally, the objective function is to choose A_{ij} to

$$Minimize \sum_i \sum_j A_{ij} [T^*_{ij} + C^*_{ij}] + \sum_j R_j \left(\sum_i A_{ij}\right) \tag{12}$$

subject to

$$P'_i \leq \sum_j A_{ij} P^*_{ij} \quad \forall i \quad (performance\ constraint) \tag{13}$$

and

$$\sum_i A_{kij} \leq \alpha_{kj} A_j \quad (supply\ constraint) \tag{14}$$

where R_j is the marginal recruiting cost of recruits in quality category j.

This objective function determines the minimum cost quality mix of recruits, given a first-term performance goal by military occupation. The performance constraint limits the performance, P'_i, allocated to occupation i to be less than or equal to a predetermined performance goal. The performance apportioned to occupation i is equal to the sum of the expected value of performance of a recruit from recruit category j in occupation i over the first-term of service, P^*_{ij}, multiplied by the number of recruits from recruit category j allotted to occupation i, A_{ij}. The supply constraint limits the number of applicants with a particular attribute related to performance (e.g., spatial composite) to a fixed proportion, α_{kj}, of a larger recruit category, A_j. In fact, the proportions are equal to the proportion of individuals in the recruiting population exhibiting that particular attribute.

Model Dimensions and Data

Dimensions

This section describes the dimension of the LP model used to evaluate alternative selection and classification models. There were nine occupational categories, corresponding to the nine one-digit Department of Defense (DoD) Enlisted Occupational Areas. Army Military Occupational Specialties (MOS) were mapped into these nine Occupational Areas, which include (a) Infantry, Gun Crews, and Seamanship Specialists; (b) Electronic Equipment Repairmen;

(c) Communications and Intelligence Specialists; (d) Health Care Specialists; (e) Other Technical and Allied Specialists; (f) Functional Support and Administration; (g) Electrical/Mechanical Equipment Repairmen; (h) Craftsmen; and (i) Service and Supply Handlers. Performance goals were specified for each occupational category. They were typically anchored by estimating the implied level of expected performance over the first-term of service supplied by an historical cohort of accessions, as predicted from the performance equations and the survival patterns of the recruit categories. In this application we used the performance implied by scoring the Fiscal Year (FY) 1990 Army recruit cohort.

The model included two major categories of recruits: "high" and "low" quality. "High quality" recruits consisted of recruits scoring in AFQT Categories I-IIIA who were high school graduates. "Low quality" recruits were those scoring in AFQT Categories IIIB and IV who were also high school diploma graduates. Within the high and low quality categories, however, there were a variable number of subcategories defined by the selection and classification models being considered. These were assumed to be available in fixed proportions within a given overall quality category, where the proportions were determined by the proportions of that subcategory in the synthetic sample of accessions within the overall category.

For example, high quality recruits consisted of recruits in AFQT Categories I, II, and IIIA. A given number of high quality recruits, N, were assumed to consist of proportions X, Y and 1-X-Y of the these three categories, respectively. A given AFQT category, such as Category I, was further divided into cells representing score ranges on other tests. The proportions, again, were determined by the proportions of the synthetic sample in those cells. In the selection and classification model with the greatest number of selection and classification tests, a total of 320 performance cells were defined.

Data

Three types of costs were used in the model: training costs, compensation costs, and recruiting costs. Training cost data for initial skill training were from the Army's Training and Doctrine Command (TRADOC) ATRM-159 report. These data were aggregated from MOS into occupational categories. The MOS level initial skill training cost data were weighted by the number of accessions in FY 1990 to form the (weighted) average training cost within an occupational category.

Compensation costs were computed for the average progressor over the first-term of service, and included basic pay, allowances, and retirement accrual. Compensation costs did not vary by occupational category in this version of the model.

The basic model was an LP, so that a constant (marginal) cost of high and low quality recruits was included in the LP. However, actual recruiting costs are non-linear. The marginal costs of high quality recruits, for example, varies with the number recruited. To account for non-linear recruiting costs within the context of an LP, we iterated between the LP and a non-linear recruiting cost function.[8] The procedure was the following. A set of starting values for the marginal costs of recruits were entered into the linear program. The recruit quality mix implied by the LP solution was then entered into the recruiting cost function, and new marginal costs were computed. These became the values for high and low quality recruiting costs in the LP, and a new LP solution was reached. Iterations between the LP and the recruiting cost function continued until convergence was achieved. Typically, this required three to five iterations.

Survival rates by occupational group and AFQT category were estimated from life tables derived from the FY 1986 cohort of Army accessions. Accessions were partitioned by occupational group and, within occupational group, by high school graduation status and AFQT status. Loss rates were computed over four years of service by counting the numbers surviving at selected intervals from the accession date to the completion of 48 months of service.

Prediction of job performance were based on regression models relating a measure of job performance to entry test scores. The details of these models are discussed below. Here, it is sufficient to note that these performance equations were used to score the synthetic sample, which contains the test scores. The number of tests within a given selection and classification model, along with how scores were categorized for these tests, define the 320 potential recruit cells or categories from which recruits were drawn and allocated to occupations.

Major Components

There are four primary components of the S&CEM: (1) the performance model; (2) survival rates; (3) recruiting, training, and compensation cost; and (4) the supply model. The model is modular in design, i.e., each of the components can be modified or replaced without affecting the others.

Performance Model

Of the several components that constitute the S&CEM, the performance model was the most important, because it described the relationship between job performance and recruit characteristics. The model's equations were used to predict the expected performance of a potential recruit in each of the occupational categories considered by the model, based on the recruit's quality characteristics and the characteristics of alternative occupations. Recall that the objective function in the model minimized costs subject to the constraint that the following performance goal was met or exceeded:

$$P'_i = \sum_j A_{ij} P^*_{ij} \qquad \forall i \tag{15}$$

where P^*_{ij} was the expected performance for a recruit from category j selected into job i over the first-term of service, and A_{ij} was the number of recruits from category j allocated to job i.

---

[8] The non-linear recruiting cost function used in our analysis is the function we developed for the Army for the Cost-Performance Tradeoff Model (McCloy et al., 1992).

Expected performance over the first-term of service was calculated for each recruit category (j) and occupation (i) as

$$P^{*}_{ij} = \sum_{t=1}^{48} S^{t}_{ij} P_{ij} \qquad (16)$$

where S^{t}_{ij} was the probability of a recruit from category j in occupation i surviving to month of service t, and P_{ij} was the expected performance from a recruit in category j and occupation i. The performance equations provided the P_{ij} term in this function.

Since the Project A data are nested, i.e., individuals are nested within jobs, the multilevel regression methodology used in the Linkage Project was adapted to estimate P_{ij}. In a multilevel model, the individual characteristics are used to predict performance, and job characteristic data are used to predict variation across jobs in the coefficients of the individual characteristics (cf. Harris et al., 1991; McCloy, Hedges, & Harris, 1991).

For the present research a fixed effects approximation to the multilevel regression approach of Harris et al. (1991) was used to provide predicted performance estimates for the decision model. The regression models have the following form:

$$P_{ij} = \alpha + \beta IC_j + \pi M_i + \rho IC_j M_i \qquad (17)$$

where P_{ij} is the performance of person j in job i; α, β, and ρ are the mean values of the regression parameters across all jobs; IC_j are the individual characteristics of person j; and M_i are the job characteristic variables for job i. The job-specific intercepts are modeled by the πM_i terms, and the job-specific slopes are modeled by the $\rho IC_j M_i$ terms (cf. Harris et al., 1991).

The presence of correlated measurement error for all individuals nested within a given job leads to incorrect standard errors when a conventional ordinary least-squares (or, "fixed effects") regression procedure is implemented. Nevertheless, the fixed effects equations can be used to approximate the multilevel models by including the job characteristic variables in the model, both as main effects and as interaction terms with the individual characteristics. Although the parameters' standard errors are incorrect (usually downwardly biased), the performance predictions typically change very little (McCloy, Hedges, & Harris, 1991).

The performance data used in the S&CEM were collected as part of Project A, a larger study of Army job performance sponsored by the Army Research Institute (Campbell, 1986, 1987; Campbell & Zook, 1990). The database is comprised of the nine Batch A MOS (and their respective measures), and the ten Batch Z MOS, for which less extensive performance data (e.g., no MOS-specific job knowledge or hands-on tests) were obtained (see Table 1).

In addition to performance measures, Project A researchers also developed several new predictor measures covering both the cognitive (e.g., written measures of spatial ability, and computerized measures of psychomotor

TABLE 1

The Military Occupational Specialties (MOS) Studied in Project A

Batch A

11B Infantryman - responsible for basic weapons, field techniques, unit tactics

13B Cannon Crewman - participates in transporting and operating field artillery equipment

19E Tank Crewman - responsible for driving tank and operating weapons system

31C Single Channel Radio Operator - operates radio, teletype, and satellite equipment

63B Light-Wheel Vehicle Mechanic - troubleshoots problems and performs regular maintenance

88M Motor Transport Operator - drives large trucks and semi-trailers

71L Administrative Specialist - performs variety of clerical and administrative tasks

91A Medical Care Specialist - administers emergency treatment and assists in outpatient and inpatient care under supervision of a physician

95B Military Police - supports battlefield operations, carries out law enforcement and security operations

Batch Z

12B Combat Engineer - assists in construction and demolition duties in the field

16S MANPADS Crewman - prepares and fires the MANPADS missile system

27E TOW/DRAGON Repairer - performs basic maintenance on TOW and DRAGON anti-tank missiles

51B Carpentry/Masonry Specialist - performs basic carpentry and masonry construction tasks

54E Chemical Operations Specialist - performs chemical reconnaissance, operates and maintains detection and decontamination equipment

55B Ammunition Specialist - assists in storage and maintenance of explosives and ammunition

67N Utility Helicopter Repairer - performs basic field and depot maintenance

76W Petroleum Supply Specialist - receives, stores, accounts for, and ships bulk and packaged petroleum supplies

76Y Unit Supply Specialist - receives, stores, accounts for, and issues all supplies for a unit.

94B Food Service Specialist - assists in the planning and preparation of meals

Note: The alphanumeric code is the Army's designation for the MOS. Batch A jobs received more extensive criterion measurement than Batch Z jobs.

---

ability and perceptual speed and accuracy), and non-cognitive (e.g., temperament, interests) domains. Table 2 lists the composite scores from the Project A predictor battery that were obtained for this project.

TABLE 2

Composite Scores From the Project A Predictor Battery

Predictor Composite:

From the ASVAB
- Technical (TCH)
  - Mechanical Comprehension
  - Auto Shop
  - Electronics Information
- Quantitative (QUN)
  - Quantitative
  - Arithmetic Reasoning
- Verbal (VRB)
  - Verbal
  - General Science
- Speed (SPD)
  - Coding Speed
  - Numerical Operations

From the Paper-and-Pencil Spatial Tests
- Spatial (SPT)

From the Computerized Perceptual/Psychomotor Tests
- Psychomotor (PSM)
- Complex Perceptual Accuracy (CPA)
- Reaction Speed (SRS)
- Complex Perceptual Speed (CPS)
- Number Speed and Accuracy (NSA)
- Reaction Accuracy (SRA)

From the ABLE
- Achievement Orientation (ACH)
- Adjustment (ADJ)
- Dependability (DEP)
- Physical Condition (CND)

From the AVOICE
- Skilled Technical (IST)
- Combat-Related (ICM)
- Food Service (IFS)
- Structural/Machines (SM)
- Audiovisual Arts (IAV)
- Protective Services (IPS)

From the JOB Questionnaire
- Organizational and Co-Worker Support (JSP)
- Routine Work (JRT)
- Job Autonomy (JAT)

The Project A criteria were used to validate both the new and extant selection measures. Table 3 details the job performance criterion measures used in the Project A concurrent validation samples. It is these data from Project A on the expanded performance and predictor domains that constitute the database for these analyses.

TABLE 3

Job Performance Criterion Measures Used in Project A Concurrent Validation Samples

---

Measures Used for all MOS

Paper-and-pencil test of Training Achievement developed for each of the 19 MOS (130-210 items each).

Five performance indicators from administrative records:

Total number of awards and letters of recommendation.
Physical fitness qualification.
Number of disciplinary infractions.
Rifle (M16) marksmanship qualification score.
Promotion rate (in deviation units).

Eleven behaviorally anchored rating scales designed to measure factors of job-specific performance (e.g., giving peer leadership and support, maintaining equipment, self-discipline).

Single scale rating of overall job performance.

Single scale rating of NCO (i.e., leadership, supervision) potential.

A 40-item summated rating scale for the assessment of expected combat performance.

Measures Used Only for Batch A MOS

From 6 to 13 MOS-specific behaviorally anchored rating scales intended to reflect job-specific technical and task proficiency.

Job sample (hands-on) measures of MOS-specific task proficiency. Individual is assessed on each of 15 major job tasks.

Paper-and-pencil job knowledge tests (150-200 items) designed to measure task-specific job knowledge on 30 major job tasks. Fifteen of the tasks were also measured hands-on.

Rating scale measures of specific task performance on the 15 tasks measured with the knowledge tests and the hands-on measures.

Situational Measures Included in Criterion Battery

A Job History Questionnaire which asks for information about frequency and recency of performance of the MOS-specific tasks.

Work Environment Description Questionnaire-a 141-item questionnaire assessing situational/environmental characteristics, leadership climate, and reward performance.

The performance criterion. The relationship between job performance and enlistment standards may be expressed as an equation in which performance is a function of some number of individual characteristics. For example, in the Linkage Project, scores on the JPM hands-on performance test were modeled as a function of a soldier's (a) AFQT score, (b) Technical composite score, (c) high school graduation status, and (d) number of months of military service (Harris et al., 1991; McCloy et al., 1992). The hands-on test provides perhaps the best measure available of one's task proficiency. Administered in a standardized setting, the hands-on test is a maximal performance measure--it is designed to assess how well an examinee can perform a particular set of tasks. In this respect, the hands-on test may be termed a "can-do" measure of performance. As with all maximal performance tests, there is the implicit assumption that each examinee is trying his or her best. That is, each examinee is believed to be maximally motivated during the test.

Although there might be a great deal of interest in an individual's maximal performance, employers and/or supervisors usually have a deeper interest in how well a person will perform on the job. That is, the question of primary interest is one concerning each person's typical performance on a day-to-day basis. Campbell, McCloy, Oppler, and Sager (1992) and McCloy, Campbell, and Cudeck (1992) have postulated and empirically tested a model of job performance determinants, arguing that the difference between maximal and typical performance measures is the degree to which motivation (defined as three choice behaviors) contributes variance to individual differences on the measures. Specifically, examinees are assumed to be maximally motivated when taking maximal performance tests; hence, motivation does not contribute to variation among test scores. Scores on typical performance measures (e.g., a supervisor's ratings of how one typically performs job tasks), however, can vary as a result of the change in the ratee's motivation across time and situations. As such, typical performance measures allow an extra dimension to be considered in addition to how well a person can perform job tasks--the tendency of the person to perform those tasks at a given level of proficiency.

From this perspective, the hands-on test can be argued to be an incomplete measure of job performance if one's interest lies primarily with an individual's typical performance. Note, however, that although typical performance is argued to be what most employers/supervisors are concerned about when they talk about performance, the most frequently used measures of typical performance (i.e., supervisor's ratings) can be fraught with difficulties. Because ratings are subjective evaluations, there is plenty of opportunity for criterion contamination (e.g., raters might give more weight than they should to a relevant performance variable or give some weight to irrelevant variables such as subgroup membership). Conversely, because ratings can assess all of the determinants of performance, there is a concomitant danger that they could be deficient (e.g., raters fail to adequately weight certain performance determinants). For example, a rater with limited opportunity to observe an individual's task performance might rely primarily on that performer's level of job knowledge when making a rating, a scenario Hunter (1986) proffered as accounting for the sizable direct effect of job knowledge on supervisor ratings in his job performance model.

A measure of total performance can be created by combining measures of can-do performance (assessing one's maximal performance) and will-do performance (assessing one's typical performance), thus considering both the proficiency of one's performance and the degree to which it is manifested on the job. Such a composite score was used as the performance criterion in this research. The components of the total performance criterion are the MOS-specific written school knowledge test score (a can-do measure) and three will-do composites: Effort and Leadership,[9] Maintaining Personal Discipline, and Physical Fitness and Military Bearing (cf. Campbell, McHenry, & Wise, 1990). The four components were standardized within job and then weighted by values of importance obtained in an earlier Project A expert judgment study[10] (Sadacca, Campbell, White, & DiFazio, 1989). The sum of these standard scores yields the composite performance criterion.

Individual characteristics. Because the criterion used here is a more expansive performance variable than the hands-on test score (i.e., it may also assess individual variation in motivation), there was reason to believe that significant additional prediction would be provided by expanding our individual characteristic variables to include non-cognitive measures. Specifically, the non-cognitive measures from Project A--in particular, scores on the temperament composites from the Assessment of Background and Life Experiences (ABLE)--have been shown to provide significant incremental validity over the cognitive measures in the prediction of the will-do criterion composites (McHenry et al., 1990). Hence, the non-cognitive measures were included as predictors of our composite performance criterion.

The individual characteristics that were examined as predictors of the performance criterion were derived from the Concurrent Validation (CV) sample from Project A (cf. Campbell, 1986), and are given in Table 2. Predictor composites were used rather than the individual scales to keep the number of independent variables at a manageable level. The cognitive predictors included (a) 4 ASVAB composites, (b) the Project A Spatial composite, which is a function of 6 paper-and-pencil spatial tests, and (c) 6 composites formed from the 20 test scores from the Project A computerized test battery of perceptual speed and psychomotor ability. The Project A non-cognitive predictors included (a) 4 temperament composites formed from 7 of the 11 substantive scales from the ABLE, (b) 6 interest composites formed from the 21 scales constituting the Army Vocational Interest Career Examination (AVOICE), and (c) 3 composites formed from the 6 scales from the Job Orientation Blank (JOB).

Correction for range restriction. The Project A data were collected on Army soldiers--individuals who had been selected into the Army on the basis of their scores on the ASVAB. For this reason, the range of scores on the ASVAB in the CV database is restricted, as are the scores on any other variables that are correlated with the ASVAB. The larger the correlation between these other variables and the ASVAB, the greater the restriction on them. Another way to say this is that there is explicit selection on the ASVAB, and incidental selection on the other variables (Lord & Novick, 1968). Both explicit and incidental selection lead to attenuation of the correlations between the restricted variables and any other variables. Such attenuation can be vexing in many applications, but its presence would be particularly damaging to an evaluation of various selection and classification models. An unbiased evaluation requires a database from a sample that has not already been selected and/or classified. In particular, we desire data for the population from which we select people.

---

[9] The Effort and Leadership (ELS) variable for the 9 Batch A MOS is different than that for the additional 10 Batch Z MOS. For the Batch A MOS, ELS contains two scores from MOS-specific behaviorally anchored rating scales. The Batch Z MOS do not include this rating scale. Rather than modifying the ELS composite by removing the MOS-specific ratings, we chose to retain these job-specific rating scales, given that their overall contribution to the final criterion composite is minimal. The same decision was not made for the can-do criteria, however, because the disparity between these criteria for Batch A (MOS-specific job knowledge, hands-on, and school knowledge tests) and Batch Z (MOS-specific school knowledge tests) was judged to be too great.

[10] The school knowledge score was weighted by the sum of the values for Core Technical Proficiency and General Soldiering Proficiency.

To correct the relationships among the observed variables for range restriction, a formula given by Lord and Novick (1968, p. 147), was applied to V, the variance-covariance matrix of the p explicit selection variables and the q incidental selection variables in the selected group (i.e., the observed variance-covariance matrix for the predictors and the criterion):

$$\boldsymbol{V} = \begin{Vmatrix} V_{p,p} & V_{p,q} \\ V_{q,p} & V_{q,q} \end{Vmatrix} . \tag{18}$$

The correction uses $W_{p,p}$ (the variance-covariance matrix of the p explicit selection variables in the unselected group) and the submatrices of **V**. For the current analyses, the population matrix $W_{p,p}$ is the variance-covariance matrix for the (p=9) ASVAB subtests from the 1980 youth population.[11] The q incidental selection variables are the additional Project A predictors (the Spatial composite score, the computerized measures, and the ABLE temperament inventory) and the composite performance criterion (q=21). Formally, the correction is

$$\boldsymbol{S}^*_{xy} = \begin{bmatrix} \boldsymbol{W}_{p,p} & \boldsymbol{W}_{p,p}\boldsymbol{V}^{-1}{}_{p,p}\boldsymbol{V}_{p,q} \\ \boldsymbol{V}_{q,p}\boldsymbol{V}^{-1}{}_{p,p}\boldsymbol{W}_{p,p} & \boldsymbol{V}_{q,q} - \boldsymbol{V}_{q,p}(\boldsymbol{V}^{-1}{}_{p,p} - \boldsymbol{V}^{-1}{}_{p,p}\boldsymbol{W}_{p,p}\boldsymbol{V}^{-1}{}_{p,p})\,\boldsymbol{V}_{p,q} \end{bmatrix} . \tag{19}$$

The resulting matrix $\mathbf{S}^*_{xy}$ contains estimates of the variances of the predictors and the criterion in the population, and of the covariances between the measures. The matrix $\mathbf{S}^*_{xy}$, which has dimensions (p+q, p+q), was scaled to a correlation matrix, $\mathbf{R}^*_{xy}$. The corrected correlation matrix $\mathbf{R}^*_{xy}$ was used for all subsequent analyses.

Survival Rates

Survival rates, describing the proportion of a given entry cohort that remains in service for particular durations over the first-term, are important because they affect expected performance and costs. The recruit must survive to "be there" to contribute to the performance of the first-term force. Moreover, recruits who leave service early are costly because this turnover implicitly generates additional training and recruiting costs.

---

[11] There are only 9 subtests here because the matrix contains the Verbal (VE) composite, which is the sum of the standardized Word Knowledge (WK) and Paragraph Comprehension (PC) subtests.

The survival rates utilized in this study were obtained from the CPTM (McCloy, et al., 1992), and predict survival over the first-term of service by recruit quality category and occupation. They are the "S^{t}_{ij}s" in the model. The survival estimates for each recruit quality category and occupational category are combined with the performance estimates to compute the performance goals, P'_{i}, for each occupational category. They also interact with costs (described below) to produce expected costs over the first-term of service, by recruit category and occupational category.

Recruiting, Training, and Compensation Costs

The recruiting cost provided an estimate of total recruiting costs as a function of the total number of recruits accessed in each quality category, the prices of recruiting resources, and recruiting environment factors, such as the unemployment rate, size of the youth population, and entry-level military pay compared to entry-level civilian pay. It was denoted by "R_j" in the mathematical statement of the model.

The recruiting costs for the two recruit quality categories were obtained indirectly from the recruiting cost function resident in the CPTM (McCloy et al., 1992). The recruiting costs in the CPTM were derived from an underlying enlistment supply curve.[12] The recruiting cost function provided the minimum cost of recruiting a given number and mix of accessions.

This module included the costs of basic and initial skill training. Basic training was constant, whereas initial skill training varied by occupational category. Compensation costs included basic pay, allowances, and retirement accrual over the first-term of service.

Supply Model

Our framework for evaluating selection and classification tests differs from some others in that the pool of recruits to be classified was endogenous.[13] Recruiting resources were increased in order to "purchase" additional higher quality recruits, or reduced to substitute less expensive lower quality recruits, in order to meet first-term performance goals at the lowest possible recruiting, training, and compensation costs. The recruit supply model is the component of the evaluation framework that permits an "endogenous" enlistment pool.

There are two notable features of this supply model. First, there are only two broad classes of recruits explicitly determined by the model--"high" quality recruits (those scoring in the upper half of the AFQT distribution and who are high school graduates) and "low" quality recruits (those scoring in AFQT categories IIIB and IV and who are high school graduates.) Yet, there are 320 performance group categories. The reason for this is that the enlistment supply literature and, indeed, an analysis of actual recruiting behavior, suggests that differential supply functions for very fine gradations of recruit quality categories are difficult to identify empirically. Second, the LP obtains its solution using a constant average cost for high and low quality recruits, respectively. That is, recruiting costs enter the LP linearly. Yet, recruiting costs are clearly non-linear. The marginal cost of additional high quality recruits increases with the numbers recruited. Below we elaborate on these two features.

---

[12] Enlistment supply curves have been a major focus of military manpower research since the institution of an all-volunteer force in 1973. They are empirically estimated equations that describe the number of personnel, by quality category, that can be recruited as a function of the recruiting environment and the quantity of recruiting resources employed.

[13] This is in contrast to a framework in which a fixed distribution of recruits are classified based on performance equations. In our framework, the quality distribution of recruits is determined along with the allocation to occupational groups. It is in this sense that the recruit distribution is endogenous.

The LP model itself recognized only two broad quality categories of recruits, "high quality" and "low quality". The proportion of high and low quality recruits obtained varied to minimize the cost of meeting performance goals. However, within the two broad categories of recruits, other subcategories were provided in fixed proportion. Hence, recruiting costs appear as:

$$\sum_j R_j A_j \tag{20}$$

in the objective function of the LP, where j varied only from 1 (high quality) to 2 (low quality). However, other recruit quality categories, up to 320 of them, were defined by the particular selection and classification tests considered in the analysis. The supply of these particular cells or categories of potential recruits was assumed to be available in fixed proportions within the high or low quality categories that they fell. Mathematically, this implied constraints on the supply of recruits of the following form:

$$\sum_i A_{kij} = \alpha_{kj} A_j \tag{21}$$

where α_{kj} was the proportion of recruits in quality category j that was in subcategory k.[14]

Recruiting costs are given by R_j--the average recruiting cost for recruits from broad quality category j. Because it is a *linear programming algorithm*, the costs are necessarily constant. However, we knew that recruiting costs for high quality recruits are inherently non-linear, increasing with the number recruited. To capture this non-linearity, we iterated between the constant costs of the linear program and a non-linear recruiting cost function.[15] We began with a set of starting values for the costs of high and low quality recruits in the linear program. The LP was exercised, and the resulting quantities of high and low quality recruits in the initial LP solution were entered into the recruiting cost function. From the recruiting cost function, we obtained a new set of (marginal) recruiting costs, to enter into the linear program. The linear program was rerun, and a new set of quantities for the recruit quality groups was determined. These were entered once again into the recruiting cost function. The process was continued until convergence was achieved. That is, when the quantities of high and low quality recruits produced by the LP solution resulted in marginal recruiting costs in the recruiting cost function that were approximately equal to the recruiting costs entered into the LP to produce those quantities, convergence was (approximately) achieved. Typically, this required three to five iterations.

---

[14] The "fixed proportion" notion is consistent with a plausible model of the actual recruiting process. Recruiters can target "high" and "low" quality applicants, perhaps, by choosing where to focus recruiting efforts. They cannot, however, target particular categories within these broader classes. An analogy to fishermen casting a net is appropriate. They can choose to effect the average mix of fish by where they fish but cannot target a specific species.

[15] The recruiting cost function used in this analysis is that developed for the Linkage Model, also called the Cost-Performance Tradeoff Model. It is derived directly from a recruit supply curve estimated using econometric methods. The recruiting cost function computes the costs of high and low quality recruits as a function of the quantities of recruits and the prices of key resources, such as recruiters, advertising, and educational incentives, while adjusting for external factors affecting recruiting costs, such as the level of unemployment, relative military and civilian pay, and the size of the youth population. See McCloy et al., (1992).

Hence, we iterated between constant recruiting costs of the linear program and the costs implied by a non-linear recruiting cost function, simulating a non-linear supply function. It is a property of linear programming solutions that if the correct marginal cost is included--the marginal cost that we would obtain at the optimal solution--the linear program will solve for the correct solution even though the marginal costs are treated as constant. Our iterative solution method takes advantage of this. While marginal recruiting costs in the LP were correct, total recruiting costs were overstated, since average recruiting costs are less than marginal recruiting costs. For this reason, we estimated recruiting costs by evaluating the recruiting cost function at the LP solution for the number of high and low quality recruits.

Summary

The S&CEM developed in this project simulates a one-stage process in which recruits are simultaneously selected and assigned. This approach is more efficient than a two-stage process in which selection and classification are sequential, independent procedures. The objective of the model is to simulate the effects of changing the type and amount of testing information available to make selection and assignment decisions. It can also be utilized to compare selection and classification procedures across different numbers and configurations of job families.

A modified multilevel regression procedure was employed to compute the job-specific performance equations. This methodology is designed to produce relatively stable estimates of performance for a large number of jobs or job families by using the total sample to estimate individual characteristics (i.e., test weights) and job characteristic data to model the differences in performance requirements across jobs. One of the major advantages of multilevel regression is that it provides the capacity to develop prediction equations for jobs that do not have criterion data, because the job characteristics based on job analysis information are substitutes for performance variables.

The S&CEM is a cost-effectiveness model. That is, alternative test batteries are evaluated in terms of the recruiting, training, and compensation costs required to select and assign recruits by AFQT category to meet *a priori* performance goals for each job. This method of measuring the utility of employment testing procedures has two main advantages over the traditional Brogden-Cronbach-Gleser approach. First, utility is measured in terms of the costs of selecting and assigning recruits, instead of the dollar value of performance, less costs, obtained from alternative batteries. Second, setting *a priori* performance goals avoids the problem of using a rational approach for establishing the value or importance of jobs, because the performance standard for each job is substituted for value.

There are two major limitations of the S&CEM. First, modifying the selection and assignment algorithm to simulate a two-stage, sequential, process could result in a case where the performance goals are met, but with some selected applicants left unassigned. This is because classification improves the predicted performance of a selected group over that of simple selection and random assignment. If recruits are selected to meet a given set of performance goals across all jobs, an efficient classification procedure will improve the performance of the group through the allocation process. The result will be that fewer recruits are needed to meet the performance goal than were selected. There are a number of ways around this problem. Two methods would be: (1) constrain the number of person years in the performance goal for classification to equal the person years that were implicitly selected to meet the performance goal; or (2) sell back or credit the total costs with the marginal cost of any selected applicants that are not required.

A second limitation of the S&CEM pilot tested in this project involves the linkage of recruiting costs to AFQT categories. It is increasing marginal costs that constrains the model from seeking only the highest quality applicants. If this were not the case the model would choose only AFQT Category I recruits. A problem will arise if a new applicant attribute is found to be related to expected performance, but for which there is no recruiting cost penalty. For example, suppose it were found that left-handers had significantly higher performance levels than right-handers. In the absence of a cost penalty or other constraint, a cost-effectiveness model like the S&CEM will choose all left-handed applicants, even though actual supply conditions are such that they would not be available. The important point is that a new applicant attribute related to performance can not be added without adding something to the recruiting cost function, or other constraint, to limit the supply of applicants with that attribute to a realistic number. Our fixed proportion assumption coupled with the recruiting cost function addresses this problem.

III. Method and Results

This chapter presents the results of the pilot test of the Selection and Classification Evaluation Model (S&CEM). The pilot test consisted of the cost-effectiveness analysis of four candidate selection and classification batteries using the one-stage selection and classification methodology. The following sections discuss our approach to generating the synthetic sample for the simulations, the prediction equations used to define the testing scenarios, the four alternative testing scenarios, and the results of the cost-effectiveness analyses.

Generation of the Synthetic Sample

To evaluate the alternative testing schemes properly, we must actually use them to select and classify individuals, analyzing the results of each application. Clearly, a correlation matrix does not contain information about specific individuals. What is needed is a synthetic sample that may be selected and classified at will. The synthetic sample must have two qualities: (1) each person in the sample must have scores on the relevant variables, and (2) the variables must have the same pattern of relationships specified by the population correlation matrix.

The second requirement might appear to make generating a synthetic sample an onerous task. Actually, such sample generation is quite simple (cf. Johnson, Zeidner, & Leaman, 1992). The goal is to obtain a factor loadings matrix **F** with dimensions of (p+q, m), where p+q is the total number of variables, m is the number of factors, and p+q=m, such that

$$R = FF' \tag{22}$$

where **R** is a correlation matrix (here, **R** is the corrected correlation matrix, R^*_{xy}). Once derived, the matrix **F** is applied to a matrix of random normal deviates (X_{rnd}) with dimensions (n, p+q), where n is the desired size of the synthetic sample. There are several ways to obtain **F**.

One way is to obtain a full principal components solution (i.e., p+q=m) of the corrected correlation matrix. In components analysis, the loadings of **F** are the weights applied to the standardized scores on the components (**P**) to reproduce the original variables:

$$X = PF'_{pca} \quad . \tag{23}$$

Because p+q=m, PF'_{pca} will perfectly reproduce the scores on the observed variables. That is, F_{pca} contains all the information about the observed scores.

Similar to the component scores in **P**, the variables in X_{rnd} have a mean of zero and standard deviation of one. To impose the population correlation structure on X_{rnd}, simply substitute it for **P** in equation 23 to yield

$$Y = X_{rnd}F'_{pca} \quad . \tag{24}$$

The resulting output matrix **Y** is a raw data matrix with dimensions (n, p+q). The correlation matrix calculated for the variables in **Y** matches the corrected population matrix, except for discrepancies caused by sampling error. In the present application, the sampling error is minimal, given the size of our synthetic sample (n = 120,000).

Table 4 contains the frequency, cumulative frequency, percentage, and cumulative percentage distribution of the absolute difference between the correlation matrix calculated for the variables in **Y** and **R** (the corrected population matrix). As can be seen, of the 435 unique pairwise correlations more than 94 percent of the differences are less than 0.0051. This indicates that the variables in the synthetic sample have essentially the same pattern of relationships specified by the corrected population correlation matrix.

Table 4

Frequency Distribution of Residuals

Range	Frequency	Cumulative Frequency	Percentage	Cumulative Percentage
.000010 - .0000509	5	5	1.149	1.149
.000051 - .0000999	9	14	2.069	3.128
.000100 - .0005099	53	67	12.184	15.408
.000510 - .0009999	71	138	16.322	31.724
.001000 - .0050999	271	409	62.299	94.023
.005100 - .0099999	26	435	5.977	100.000

Although the principal components solution will provide the desired results, this is not the approach used by Johnson, Zeidner, and their colleagues. Rather, they use a loading matrix (call this $\mathbf{F}_{gfa}$) from what they term a "Gramian factor solution" (Johnson, Zeidner, & Leaman, 1992, p. F-1). Their approach was adopted here. The matrix $\mathbf{F}_{gfa}$ is obtained by using the eigenvectors and eigenvalues of the corrected correlation matrix in a way that differs slightly from components analysis.

Consider the components loading matrix $\mathbf{F}_{pca}$ first. In terms of eigenvectors and eigenvalues, this matrix is calculated thusly:

$$F_{pca} = W\Lambda^{1/2} \qquad (25)$$

where W is a matrix of the eigenvectors of **R** and Λ is a diagonal matrix of eigenvalues. To obtain the **F** for the Gramian factor solution, one simply postmultiplies $\mathbf{F}_{pca}$ by W':

$$F_{gfa} = F_{pca} W' = (W\Lambda^{1/2}) W' \quad . \tag{26}$$

The principal components are uncorrelated linear composites of the observed variables that account for the maximal amount of variance in the observed variables. The variance of the first component is always the largest, with each subsequent component decreasing in variance. Hence, the score variance of the observed variables is not evenly distributed throughout $\mathbf{F}_{pca}$. The sum of the squared loadings in the kth column of $\mathbf{F}_{pca}$ equals the kth eigenvalue of $\mathbf{R}$. The eigenvalues give the variances of the components, and the eigenvalues decrease (i.e., $\lambda_1 > \lambda_2 > \ldots > \lambda_k$). Hence, the majority of the score variance in the observed variables appears on the left side of $\mathbf{F}_{pca}$.

The approach of Johnson and Zeidner redistributes this variance throughout the matrix, much like gently shaking a box containing a small amount of sand from side to side more evenly distributes the sand along the box bottom. The postmultiplication shown in equation 26 standardizes the variance of each component, transforming $\mathbf{F}_{pca}$ such that the sum of the squared loadings for a component equals one for all of the components. $\mathbf{F}_{gfa}$ is also symmetric.

The postmultiplication of $\mathbf{F}_{pca}$ by W' is actually nothing more than an orthogonal transformation (i.e., rotation) of the components factor loading matrix, $\mathbf{F}_{pca}$. In general, a factor loading matrix is transformed by postmultiplying it by a transformation matrix, $\mathbf{T}$, to yield a new loading matrix, $\mathbf{F}^*$:

$$F^* = F T \quad . \tag{27}$$

If $\mathbf{T}$ is an orthonormal matrix, then $\mathbf{TT}' = \mathbf{T}'\mathbf{T} = \mathbf{I}$ (an identity matrix), and the transformation is an orthogonal rotation (i.e., the factors are uncorrelated). Note that the eigenvectors of a symmetric matrix are mutually orthogonal, with $\mathbf{W}'\mathbf{W} = \mathbf{I}$. This is true even if $\mathbf{W}$ has dimensions $(p+q, m)$ with $m < p+q$, thereby containing only the first m eigenvectors of the symmetric matrix. When $m = p+q$, then $WW' = I$, as well. As mentioned, $m = p+q$ in a full principal components solution. Hence, the matrix W from a full components solution is orthonormal. Letting $\mathbf{F} = W\Lambda^{\frac{1}{2}}$ and $\mathbf{T} = W'$ in equation 27 yields

$$F^* = F T = (W\Lambda^{1/2}) W' = F_{pca} W' = F_{gfa} \quad . \tag{28}$$

Prediction Equations

A number of regression models were estimated using the corrected population correlation matrix. These regression models related the performance criterion, P_{ij} (the criterion score of individual j in job i), to different sets of the individual characteristics. The goal was to obtain a sample of equations that depicted the effects of additional testing information (i.e., individual characteristics) on a one-stage selection and classification model. The regression equations were examined in terms of the amount of variance they accounted for in the performance criterion. The predictors used in the equations were the following (see Table 2):

-- A_s (the nine ASVAB subtests)

-- A_c (the four ASVAB composites)
-- Sp (the Spatial composite)
-- C (the six computer composites)
-- Ab (the four ABLE composites)
-- Av (the six AVOICE composites)
-- Jo (the three JOB composites)
-- FS (the four factor scores of job characteristics from the Linkage Project)

For all the equations reported below, the four factor scores appear (1) as main effects and (2) in interaction terms with each of the individual characteristics in the model. This procedure provides the main effects approximation to the multilevel model that is more appropriate for nested data (cf. Hogan, McCloy, Harris, & McWhite, 1993). In terms of the individual characteristics, the following equations were estimated:

1. $P_{ij} = A_s$
2. $P_{ij} = A_c$
3. $P_{ij} = A_s$, Sp
4. $P_{ij} = A_c$, Sp
5. $P_{ij} = A_s$, Sp, C
6. $P_{ij} = A_c$, Sp, C
7. $P_{ij} = A_s$, Sp, C, Ab
8. $P_{ij} = A_c$, Sp, C, Ab
9. $P_{ij} = A_s$, Sp, C, Ab, Av
10. $P_{ij} = A_c$, Sp, C, Ab, Av
11. $P_{ij} = A_s$, Sp, C, Ab, Av, Jo
12. $P_{ij} = A_c$, Sp, C, Ab, Av, Jo.

All the equations were estimated on the total sample (i.e., there are no job-specific equations; the factor scores provide for job-specific variation in the regression coefficients). The multiple correlations and R^2 values for these equations are given in Table 5. The results suggest the following:

(1) The use of all the ASVAB subtests singly outperforms the use of the ASVAB composites, but only slightly. Given the additional degrees of freedom they consume, the subtests were removed from further consideration.

(2) The Spatial composite provides a small but significant portion of incremental validity over the ASVAB.

(3) The Computer composites do not yield any incremental validity over the paper-and-pencil cognitive measures. (Note, however, that later results suggest that the computerized measures might provide incremental validity in a particular occupational code.)

(4) The ABLE composites yield the largest amount of incremental validity, boosting the R^2 values several percentage points.

(5) Neither the AVOICE nor the JOB composites provide incremental validity over the cognitive and ABLE measures.

TABLE 5

Multiple Correlations for 12 Alternative Equations for Predicting Performance

Equation	R	R^2
1. $P_{ij} = A_s$	.523	.273
2. $P_{ij} = A_c$	.520	.271
3. $P_{ij} = A_s$, Sp	.531	.282
4. $P_{ij} = A_c$, Sp	.527	.278
5. $P_{ij} = A_s$, Sp, C	.536	.287
6. $P_{ij} = A_c$, Sp, C	.532	.283
7. $P_{ij} = A_s$, Sp, C, Ab	.586	.344
8. $P_{ij} = A_c$, Sp, C, Ab	.583	.340
9. $P_{ij} = A_s$, Sp, C, Ab, Av	.593	.351
10. $P_{ij} = A_c$, Sp, C, Ab, Av	.590	.348
11. $P_{ij} = A_s$, Sp, C, Ab, Av, Jo	.595	.354
12. $P_{ij} = A_c$, Sp, C, Ab, Av, Jo	.592	.351

In summary, the new cognitive measures (the spatial and computer tests) provide minimal incremental validity to the ASVAB, although they exhibit respectable correlations with the criterion alone. Note that the Spatial tests could be incorporated into recruit testing with relatively little cost, given that they are also paper-and-pencil tests. In addition, the computerized measures are likely at a disadvantage with the present criterion. They are likely to be most predictive of measures that allow variation in skill and procedural knowledge, such as hands-on tests (McCloy, 1990; Campbell, McCloy, Oppler, & Sager, 1992). The incremental validity that is witnessed is provided by the ABLE composites. As reported in the previous Project A research, this set of composites shows substantial incremental validity over the cognitive measures when a criterion containing will-do measures is being used. It must be kept in mind, however, that the potential incremental validity for the ABLE may not be realized to the degree suggested by the Current Validity (CV) sample estimates. In addition to the fact that ABLE validity coefficients were lower within the Longitudinal Validity (LV) sample, research suggests that temperament measures are susceptible to faking (Young, White, & Oppler, 1991). In light of such stability and distortion issues, the findings regarding ABLE may be overly optimistic.

Based on the regression results just presented, five batteries were selected for investigation. Specifically, those individual characteristic composites with significant partial regression coefficients were selected to form a second class of regression models. In addition, we opted to split the ASVAB into two pieces: the AFQT (described in terms of composites as QUN and VRB), and the remaining subtests (described by the TCH and SPD composites).

This allowed us to examine a test battery matching the current selection procedure (Battery A below). Five batteries were chosen:

Battery A: AFQT
Battery B: AFQT, TCH, SPD
Battery C: AFQT, TCH, SPD, SPT
Battery D: AFQT, TCH, SPD, SPT, NSA, PSM
Battery E: AFQT, TCH, SPD, SPT, NSA, PSM, CND, DEP, ACH.

Again, the five test batteries also included job characteristic factor scores as main effects and as interactions with the individual characteristics. The multiple correlations and R^2 values for these equations are given in Table 6. Because little effect was observed for the computerized composites NSA and PSM, batteries D and E were collapsed to form the following:

Battery F: AFQT, TCH, SPD, SPT, CND, DEP, ACH .

This battery has a multiple correlation of .592 and R^2 of .350. Based on these findings, four test batteries were selected for the S&CEM: A, B, C, and F.

TABLE 6

Multiple Correlations for Six Alternative Selection and Classification Batteries

	Battery	R	R^2
A.	P_{ij} = AFQT	.500	.250
B.	P_{ij} = AFQT TCH SPD	.526	.277
C.	P_{ij} = AFQT TCH SPD SPT	.536	.287
D.	P_{ij} = AFQT TCH SPD SPT NSA PSM	.538	.290
E.	P_{ij} = AFQT TCH SPD SPT NSA PSM CND DEP ACH	.594	.353
F.	P_{ij} = AFQT TCH SPD SPT CND DEP ACH	.592	.350

Each of the four test batteries formed the basis of a separate personnel enlistment testing condition. We estimated the recruiting, training, and compensation costs of meeting a given set of performance goals by occupational area for each of the selection and classification batteries, A, B, C, and F. The selection and classification composites included in each of the test batteries is summarized in Table 7. The Battery F equations were used to project the performance of an historical cohort of recruits. In this instance, the FY 1990 recruit cohort was used. These projections formed the performance goals, by occupational group, which were held constant throughout the analysis.

TABLE 7

Selection and Classification Batteries Considered

Battery	AFQT	ASVAB Composites	Spatial	Able	R^2
A	X				0.250
B	X	X			0.277
C	X	X	X		0.287
F	X	X	X	X	0.350

The LP was exercised using weights from the Model F performance equations for each of the other models. However, the predictors not included in the models were evaluated at their conditional means. That is, in Model A an applicants actual AFQT score was used in predicting expected performance, but the ASVAB composites, spatial test and ABLE were evaluated at the mean, conditional on the applicant's AFQT score, not the actual scores for the applicant. Similarly, Model B used the applicants' actual AFQT and ASVAB composite scores, but spatial and ABLE variables were entered at their means, conditional on the applicants' AFQT and ASVAB composite scores. In Model F, of course, all tests were evaluated at the individual's actual scores for those tests.

By comparing the differences in recruiting, training, and compensation costs of meeting the same performance goals, we obtained dollar denominated estimates of the value of additional selection and classification information provided by each of the test batteries. The value of the tests that are included in model Y that are not in model X, then, is given by:

$$\text{Value of incremental tests} = C(X) - C(Y)$$

where C(X) represents the dollar cost of meeting performance goals given the selection and classification information contained in X.

Results

The results for each of the batteries are shown in Table 8. For Battery A, the LP suggests that a recruiting cohort of 76,971 recruits and a high quality mix of 93 percent is the lowest cost way of meeting performance goals. When additional ASVAB composites are added to the information available for making selection and classification decisions, almost 2,000 fewer accessions are required, but the high quality mix of these accessions rises to almost 96 percent. Total costs of meeting performance goals decline from $7,235.5 million to $7,086.1 million, implying that the value of the ASVAB composite information contained in Battery B is almost $150 million over the first-term of service, for this Army cohort of recruits.

Battery C adds a spatial composite to the ASVAB battery. The spatial composite apparently provides information that increases the relative value of some lower quality recruits, because the high quality mix declines to about 89

TABLE 8

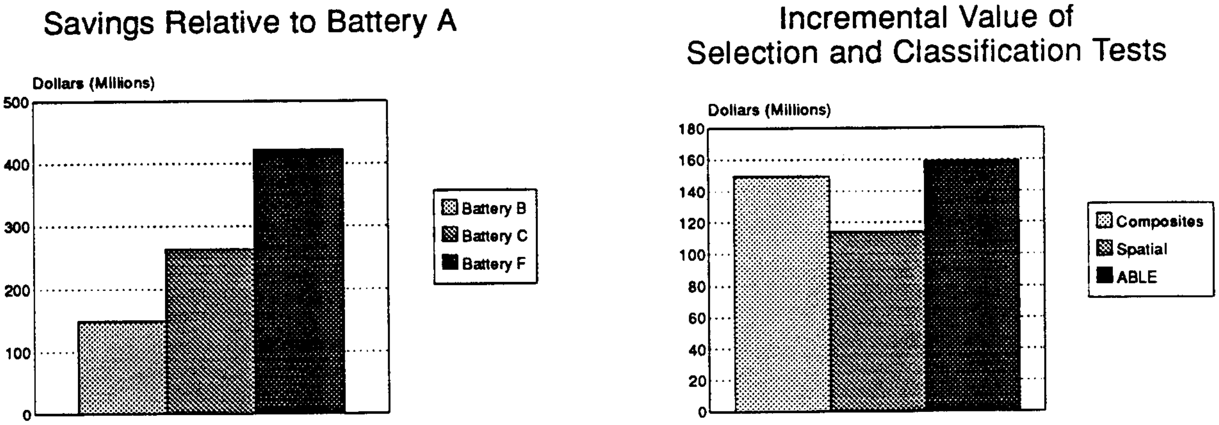
Results from Selection/Classification Battery Evaluation

Battery	Total Costs	Recruiting Costs	Training and Compensation Costs	Accessions	Percent High Quality
A	$7,235.5M	$962.7M	$6,272.8M	76,971	93.0
B	$7,086.1M	$961.8M	$6,124.3M	75,110	95.5
C	$6,972.0M	$845.5M	$6,126.5M	75,352	89.2
F	$6,812.6M	$936.8M	$5,875.8M	71,956	98.9

percent. Accessions increase only marginally relative to the Battery B solution. The incremental value of the spatial composite, in meeting first-term performance goals, is about $114 million. Finally, the ABLE test battery is added in F. Total accession requirements decline by about 3,400 relative to Battery C, and by about 5,000 relative to Battery A. The quality mix rises to its highest level, almost 99 percent, suggesting that those who are willing to work hard, as indicated by ABLE, tend also to be the most capable recruits, as measured by traditional aptitude tests. Adding ABLE to those tests included in C reduces the total costs of meeting performance goals over the first-term of service by about $155 million--a measure of the incremental value of the ABLE tests provided by this evaluation framework.

Note that most of the savings from the additional selection and classification tests is in training and compensation costs. Recruiting costs decline only modestly for Batteries B and F, relative to Battery A. Only in C, when the spatial test is initially added, is there a significant reduction in recruiting costs.

The following two charts indicate (a) the savings associated with each model, relative to Battery A and (b) the incremental value of the added tests, as implied by our cost-effectiveness framework.



Savings Relative to Battery A

Incremental Value of Selection and Classification Tests

Figure 1. **Cost-effectiveness comparisons of personnel selection and classification tests**

IV. Discussion and Conclusions

This project had two objectives. The primary goal was to develop a new methodological framework for evaluating selection and classification procedures. The second objective was to examine the practical utility of the model by utilizing it to evaluate the efficiency of a small number of alternative test batteries.

Unlike previous efforts in this area, the Selection and Classification Evaluation Model (S&CEM) developed in this study utilizes a cost-effectiveness framework based on a cost-minimization strategy. This methodology places a dollar value on the changes in recruiting, compensation, and training resources that would occur with incremental changes in test batteries and other components of a selection and classification system. The evaluation framework improves upon previous research in three major ways.

First, the S&CEM can simulate three alternative selection and classification processes: (1) single-stage, (2) multi-stage, and (3) simultaneous. Previous evaluation models either provided classification only of a preselected applicant group across multiple jobs, or selection only into a single job. The procedure examined in this study was single-stage simultaneous selection and classification, where the number and quality mix of recruits was determined within the model (as part of the cost minimization process), according to the enlistment supply and demand relationships that are part of the model.

Second, the S&CEM estimates the value of incremental tests in a dollar metric that can be directly related to programs and budgets. In the cost-effectiveness framework, the value of the selection and classification information provided by incremental tests is measured by the difference in recruiting, training, and compensation costs that must be incurred to meet the performance goals associated with a particular recruit cohort. In the past, the benefits of improved selection and/or classification were measured either in terms of the physical units associated with performance measurement, which begged the question of the "value" of the increased performance, or by a somewhat ephemeral measure of the dollar value of the increased "utility" provided by improved performance.

Third, the model places the estimation of the value of selection and classification tests within a coherent framework of the recruiting and training personnel required to meet readiness or performance goals. For the first time, an evaluation framework considers all, or most, of the key factors that should affect the costs of meeting readiness goals for the first-term force, including:

(1) Marginal recruiting costs that differ by quality characteristics and increase as more are recruited;

(2) Training costs that vary by occupational field;

(3) Attrition rates that vary with recruit quality characteristics; and

(4) Expected performance that changes both among recruits of varying individual characteristics and among occupations for a given recruit.

An application of the S&CEM was conducted in the second phase of this study to investigate the model's potential usefulness to the Army for evaluating selection and classification procedures. A single-stage process was simulated using a linear programming model in which synthetic recruits were simultaneously screened and assigned to one of nine occupational areas to meet performance goals in those areas at the lowest recruiting, training, and compensation costs.

Four test batteries, which increased in the number and dimensionality of the predictors, were evaluated. Battery A contained only the Armed Forces Qualification Test (AFQT). Battery B added the verbal, quantitative, technical, and speed composites of the Armed Services Vocational Aptitude Battery (ASVAB). Battery C added the Project A spatial composite, and Battery F added the ABLE, a measure of motivation to perform. Each battery was used to select and classify a synthetic recruit cohort to produce specific levels of predicted performance in nine occupational areas over four years of service. The synthetic recruits and the performance standards approximated the Fiscal Year (FY) 1990 recruit cohort and their predicted levels of job performance.

Three major conclusions were derived from the results of the pilot test of the S&CEM. First, adding a spatial composite to the ASVAB (Battery C) could save the Army up to $114 million in recruiting, training, and compensation costs for a recruit cohort over four years. Interestingly, the spatial composite seems particularly useful in finding occupational areas where lower quality recruits, as measured by AFQT score, with above average spatial ability would perform well.

Second, adding the ABLE to the ASVAB and a spatial composite (Battery F) resulted in estimated savings of $160 million relative to Battery C, and the selection of a higher quality mix of recruits. The latter finding is due to the high correlation of some ABLE scales with the cognitive predictors. Lastly, a comparison across all four test batteries showed that adding tests of new cognitive and noncognitive factors to the ASVAB composites improved selection and classification decisions by meeting performance goals at lower costs.

Several limitations in the pilot study analyses were noted. First, although the linear programming method provided a relatively clear answer to the question of the value of better selection and classification methods, this approach assumes an "optimal" selection and classification of recruits based on expected performance and costs. It does not explicitly consider factors such as applicant preferences and/or training seat availability that may limit the extent to which Army counselors could "optimally" classify recruits in practice. Hence, to the extent that the additional selection and classification information is used less than "optimally," as defined here, the values placed on improved selection and classification methods may be overstated.[16] Future research, within this cost-effectiveness framework, could more closely attempt to model how the information would be used in practice by Army counselors.

---

[16] On the other hand, since we estimate the value as differences from a base case, one might argue that any overstatement due to the assumption of "optimal" use of the selection and classification information is impounded in all cases, and that this effect is "differenced out" in the comparisons.

Second, only two broad categories of recruits were included in the supply analysis. Moreover, non-high school graduates were excluded in order to make the number of performance cells (320) tractable. Future research should attempt to expand the number of recruit supply categories that are explicitly modeled, and to include non-graduates in the analysis.

Third, we did not explicitly consider the costs associated with generating better selection and classification information. Rough estimates indicate that these costs would reduce the incremental value of selection and classification tests only marginally.[17]

Finally, our analysis was done in a risk neutral, expected value framework. Improved models of selection and classification undoubtedly increase the precision with which performance is forecast. If the Army is risk aversive, then the value of improved selection and classification methods is understated using our framework. Future efforts could incorporate the value of improved precision within our overall framework.

All in all, the savings estimated for alternative selection and classification models should be considered as relative rather than absolute values. The S&CEM is a useful analytic tool for assessing the potential value of additional tests developed in non-operational contexts. The evaluation framework developed here can be applied to a number of different policy issues facing the Army. Examples of some specific policy questions and issues that may be evaluated with the current framework include:

(1) How would results change if we include more realistic factors, such as applicant preferences and training seat availability, directly in the simulations? What is the value (cost) of limiting (expanding) applicant choices in classification?

(2) What are the expected costs associated with eliminating a test, such as Numerical Operations, from the current selection and classification battery?

(3) What is the "optimal" set of questions to include in an aptitude test? Can an "optimal" battery be constructed using the framework?

(4) What is the dollar value of the tradeoff between tests with less adverse impact, but less predictive precision?

In summary, the Selection and Classification Evaluation Model was developed and pilot tested in this project. The results indicate that a cost-effectiveness method of evaluating selection and classification procedures is a useful research and development tool. Future applications of the S&CEM can be directed toward both expanding the analysis of test batteries and other components of the Army's selection and classification system and modeling alternative management policies and environmental factors.

---

[17] An *upper-bound* measure of the additional operating costs of adding tests to the current battery is about $2 million for Army.

REFERENCES

Brogden, H. E. (1946). On the interpretation of the correlation coefficient as a measure of predictive efficiency. Journal of Educational Psychology, 37, 65-76.

Brogden, H. E. (1949). When testing pays off. Personnel Psychology, 2, 171-183.

Brogden, H. E. (1951). Increased efficiency of selection resulting from replacement of a single predictor with several differential predictors. Educational and Psychological Measurement, 11, 173-196.

Campbell, J. P. (Ed.) (1986). Improving the selection, classification, and utilization of Army enlisted personnel: Annual report, 1985 fiscal year (ARI Technical Report 746). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. (AD A193 343)

Campbell, J. P. (1987). Validation analysis for new predictors (ARI Working Paper 86-09). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Campbell, J. P., McCloy, R. A., Oppler, S. H., & Sager, C. E. (1992). A theory of performance. In N. Schmitt, W. C. Borman & Associates (Eds.), Personnel selection in organizations (pp. 35-70). San Francisco, CA: Jossey-Bass.

Campbell, J. P., McHenry, J. J., & Wise, L. L. (1990). Modeling job performance in a population of jobs. Personnel Psychology, 43(2), 313-333.

Campbell, J. P., & Zook, L. M. (1990). Improving the selection, classification, and utilization of Army enlisted personnel: Final report on Project A (ARI Research Report 1597). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. (AD A242 921)

Cronbach, L. J., & Gleser, G. C. (1965). Psychological tests and personnel decisions (2nd ed.). Urbana, IL: University of Illinois Press.

Department of Defense. (1980). Aptitude testing of recruits. Washington, DC: Office of the Assistant Secretary of Defense (Manpower, Reserve Affairs, and Logistics).

Fernandez, R. L., & Garfinkle, J. B. (1985). Setting enlistment standards and matching recruits to jobs using job performance criteria (R-3067-MIL). Santa Monica, CA: Rand Corporation.

Harris, D. A., McCloy, R. A., Dempsey, J. R., Roth, C., Sackett, P. R., Hedges, L. V., Smith, D. A., & Hogan, P. F. (1991). Determining the relationship between recruit characteristics and job performance: A methodology and a model (HumRRO Report FR-PRD-90-17). Alexandria, VA: Human Resources Research Organization.

Hogan, P. F., McCloy, R. A., Harris, D. A., & McWhite, P. (1993). Estimating the benefits of selection and classification models (HumRRO Report IR-PRD-93-02). Alexandria, VA: Human Resources Research Organization.

Hunter, J. E. (1986). Cognitive ability, cognitive aptitudes, job knowledge, and job performance. Journal of Vocational Behavior, 27, 340-362.

Hunter, J. E., & Schmidt, F. L. (1982). Fitting people to jobs: The impact of personnel selection on national productivity. In E. A. Fleishman & M. D. Dunnette (Eds.), Human performance and productivity: Vol. 1: Human capability assessment. Hillsdale, NJ: Erlbaum.

Johnson, C. D., Zeidner, J., & Leaman, J. A. (1992). Improving classification efficiency by restructuring Army job families (ARI Technical Report 947). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. (AD A250 139)

Laurence, J. H., & Hoffman, R. G. (1993). A description and evaluation of selection and classification models (HumRRO Report FR-PRD-93-01). Alexandria, VA: Human Resources Research Organization.

Lord, F. M., & Novick, M. R. (1968). Statistical theories of mental test scores. Reading, MA: Addison-Wesley.

McCloy, R. A. (1990). A new model of job performance: An integration of measurement, prediction, and theory. Unpublished doctoral dissertation, University of Minnesota, Minneapolis.

McCloy, R. A., Hedges, L. V., & Harris, D. A. (1991). Development of a methodology to link recruit quality requirements to job performance: Estimation of model parameters (HumRRO Report IR-PRD-91-07). Alexandria, VA: Human Resources Research Organization.

McCloy, R. A., Campbell, J. P., & Cudeck, R. (1992). A confirmatory test of a model of performance determinants. Manuscript submitted for publication.

McCloy, R. A., Harris, D. A., Barnes, J. D., Hogan, P. F., Smith, D. A., Clifton, D., & Sola, M. (1992). Accession quality, job performance, and cost: A cost-performance tradeoff model (HumRRO Report FR-PRD-92-11). Alexandria, VA: Human Resources Research Organization.

McHenry, J. J., Hough, L. M., Toquam, J. L., Hanson, M. A., & Ashworth, S. (1990). Project A validity results: The relationship between predictor and criterion domains. Personnel Psychology, 43, (2), 335-354.

Nord, R. D., & Kearl, C. E. (1990). Estimating cost effective recruiting missions: A profit maximizing approach. Paper presented to the Annual Meeting of ORSA/TIMS, Philadelphia, PA.

Nord, R. D., & Schmitz, E. (1989). Estimating performance and utility effects of alternative selection and classification policies. In J. Zeidner & C. Johnson (Eds.), The economic benefits of predicting job performance (IDA Paper P-2241). Alexandria, VA: Institute for Defense Analyses.

Sadacca, R., Campbell, J. P., White, L. A., & DiFazio, A. S. (1989). Project A: Weighting criterion components to develop composite measures of job performance (ARI Technical Report 838). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. (AD A210 357)

Young, M. C., White, L. A., & Oppler, S. H. (1991). Coaching effects on the Assessment of Background and Life Experiences (ABLE). Paper presented at the 33rd Annual Conference of the Military Testing Association, San Antonio, TX.

Zeidner, J., & Johnson, C. D. (1989). The economic benefits of predicting job performance (IDA Paper P-2241). Alexandria, VA: Institute for Defense Analyses.